USER MANUAL

Data Analytics Support Tools for Clinical Risk Assessments

Abstract

Our product is a support tool, which is expected to extract information from CSV files provided by the client, re-store the selected data into new documents in required formats which will be suitable for analytical software such as WEKA data mining software and See5. The output of our product is new generated files (.arff / .names / .data) that could be used directly in analytics tools.

Yi He, Siva Polineni, Varun Yenumula
yihe@students.federation.edu

Table of Contents

Installment

Easy access to the product

For users using the windows 10 operating system, we provide an execution file (i.e. “supporttool.exe”) as a simple operation product. Users only need to download the execution file and run it in the windows system to use our tool products. No need to install additional specific software and modules.

Traditional installation approach

Users or other specialists could also access to our product via traditional installation approach – via python software opening our python file (i.e. “supporttool.py”). To make the tool work properly, you need install the following software and modules to prepare the environment.

Prerequisite installment software: Python 3.7 (64-bit), Pycharm 2018 edu.

Third-party modules needed to be installed: Pandas1.0.4 (cmd: pip install pandas).

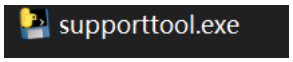
Relevant files needed to be stored in the same folder of “supporttool.py”: “convertcsvtoarff.py” and “convertcsvtosee5.py”.

Once the environment is setting as prerequisite, you can run “supporttool.py” to get access to our product. The product operation instructions will be explained in detail in next section.

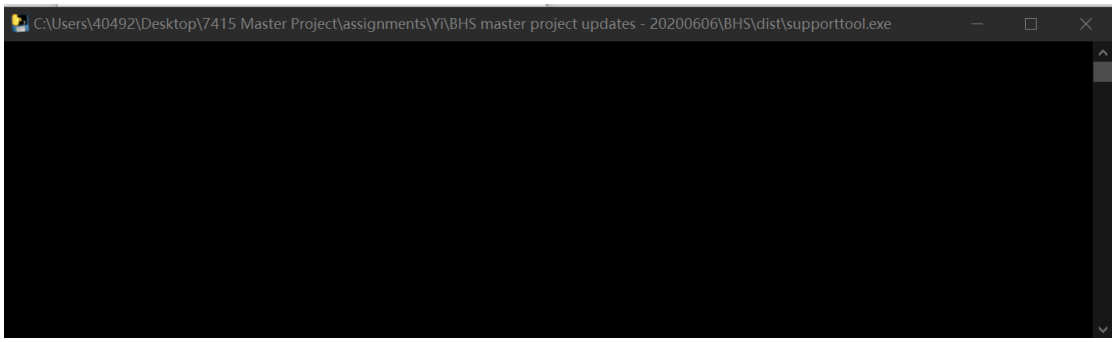
User guide

Get start with this tool

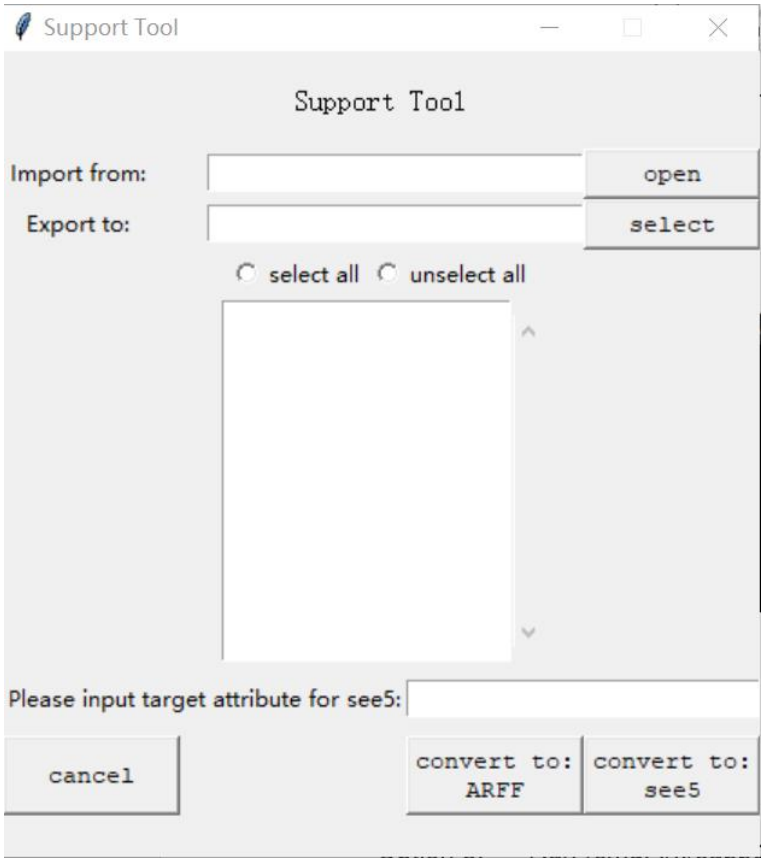
Option 1. Run (double click) the execution file “supporttool.exe”.



Wait for a few seconds.



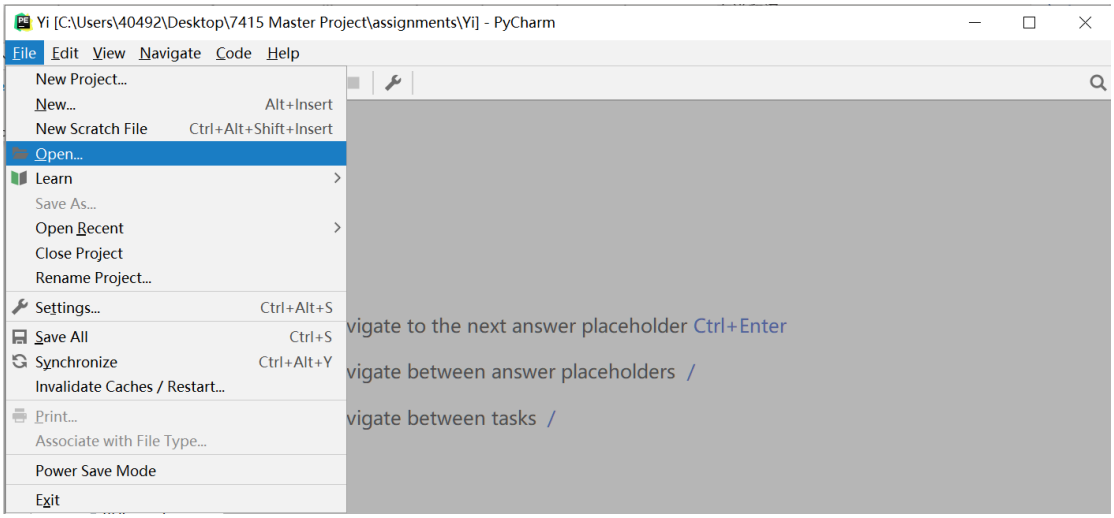
And you will access the user interface of our product.

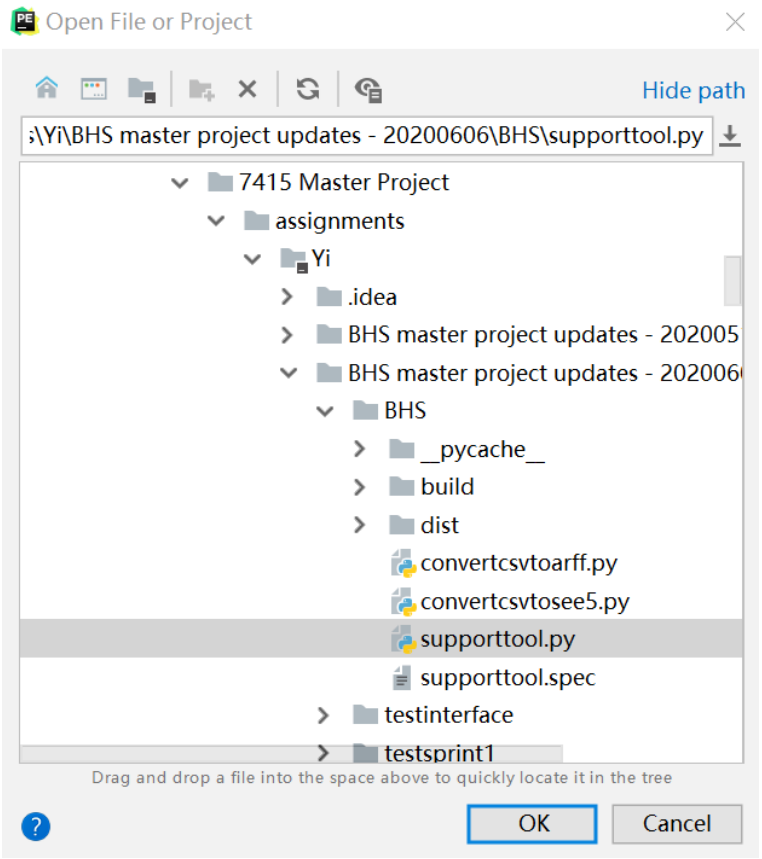


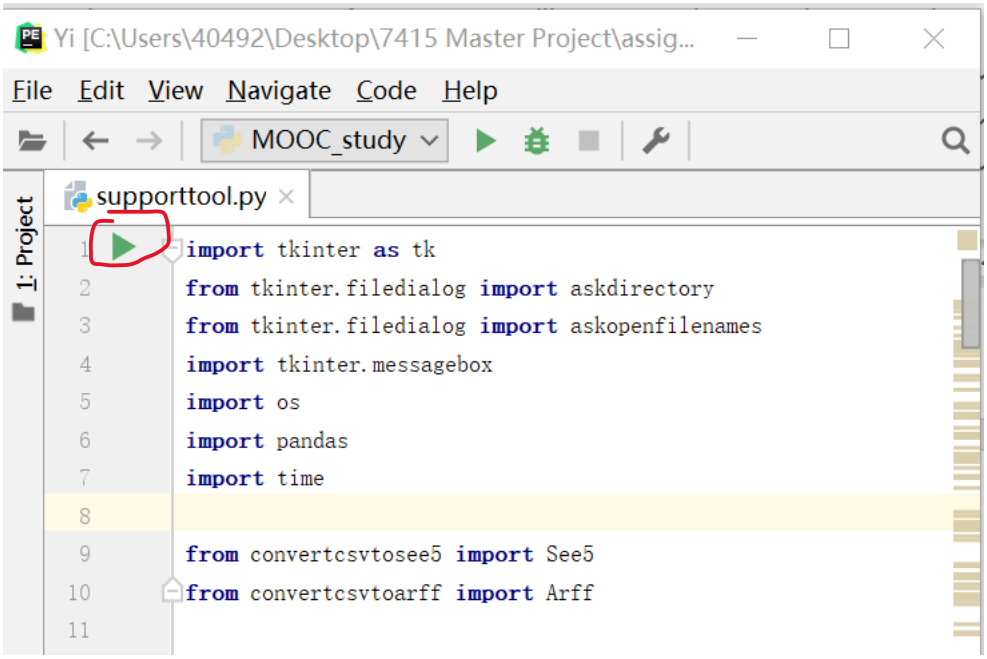
Option 2. Run (right click) the execution file “supporttool.exe” icon, select “open”.

Option 3. Run (cmd: ./../supporttool.exe) the execution file “supporttool.py”.

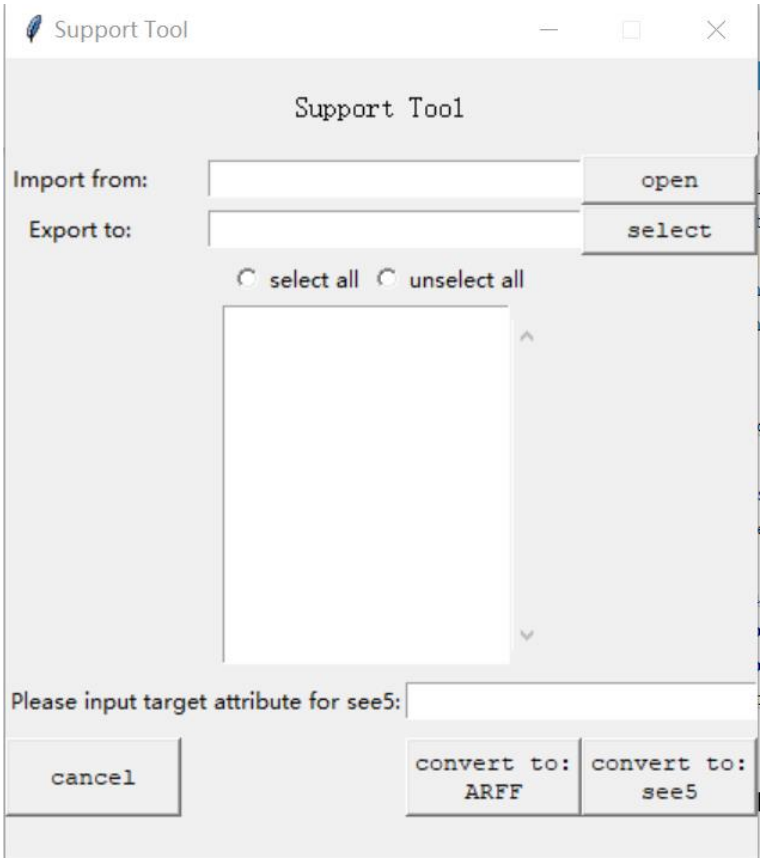
Option 4. Run Pycharm and open ‘supporttool.py”.





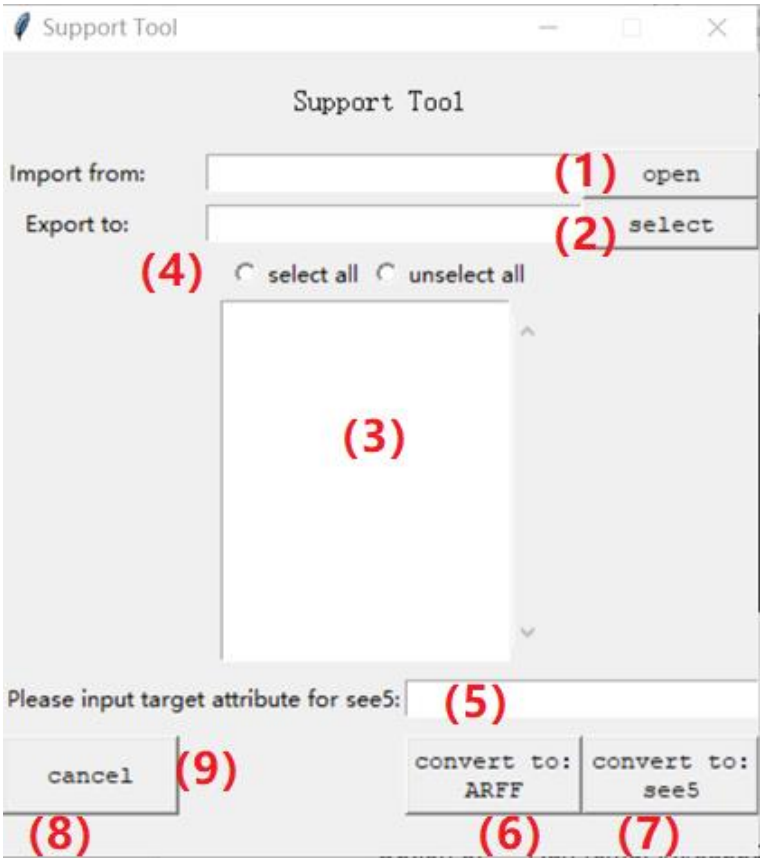


Click the green triangle on the top left (line 1 of the codes). And then the user interface will pop out automatically.



One file conversion

1. Functions of each button on the user interface.



There are 9 sections that you might use:

(1). Open button: open the csv files you want to convert in the directory where you store them, pick the csv files you want to convert and click “open”.

(2). Select button: choose the directory where you want to save your converted file.

(3). Attributes presentation area: once the csv files is opened, the attributes (first row of the csv files) will be presented in the listbox, and you can select or unselect the attributes that you want to convert from the csv file.

(4). Attributes selection options: select all – select all the attributes in the listbox, unselect all – unselect all the attributes in the listbox.

(5). Users could input the target attribute for the see5 files. If did not input anything, the target attribute will be null.

(6). Convert to ARFF format button: once you have done above steps (except step (5)), you can get the converted ARFF file (.arff) in the location you select.

(7). Convert to see5 format button: same as “convert to ARFF” format button, you can get the converted attribute file (.names) and data file (.data) in the location you select.

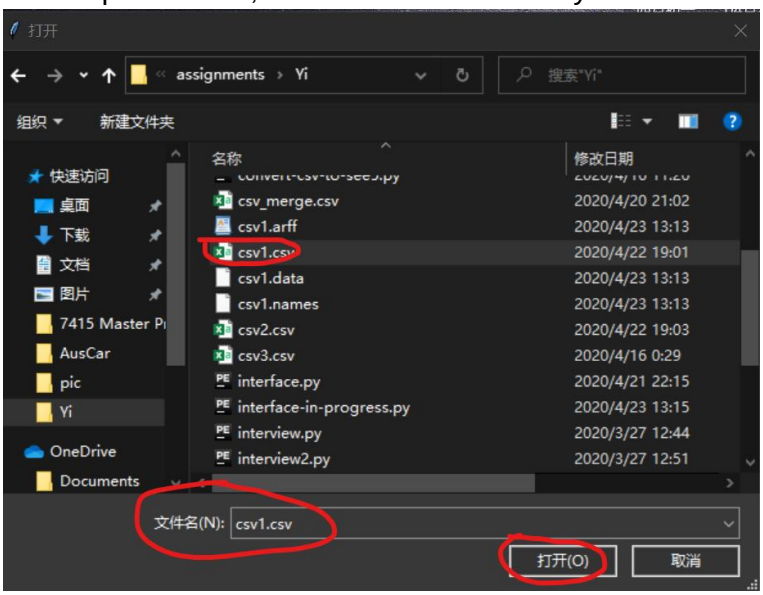
(8). Status bar: show the status of the conversion progress.

(9). Cancel button: quit this convert and close the window.

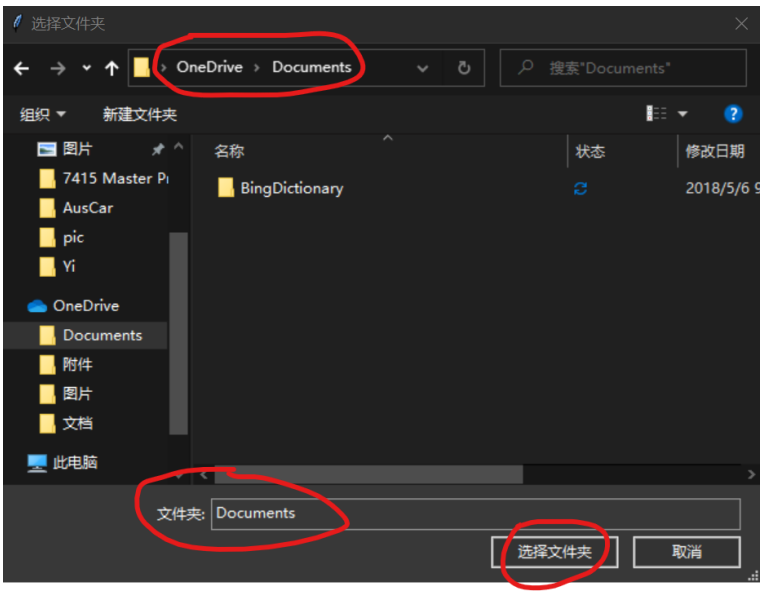
2. How to use our product.

(1). Run our product. You will see a window as the initial interface picture shows.

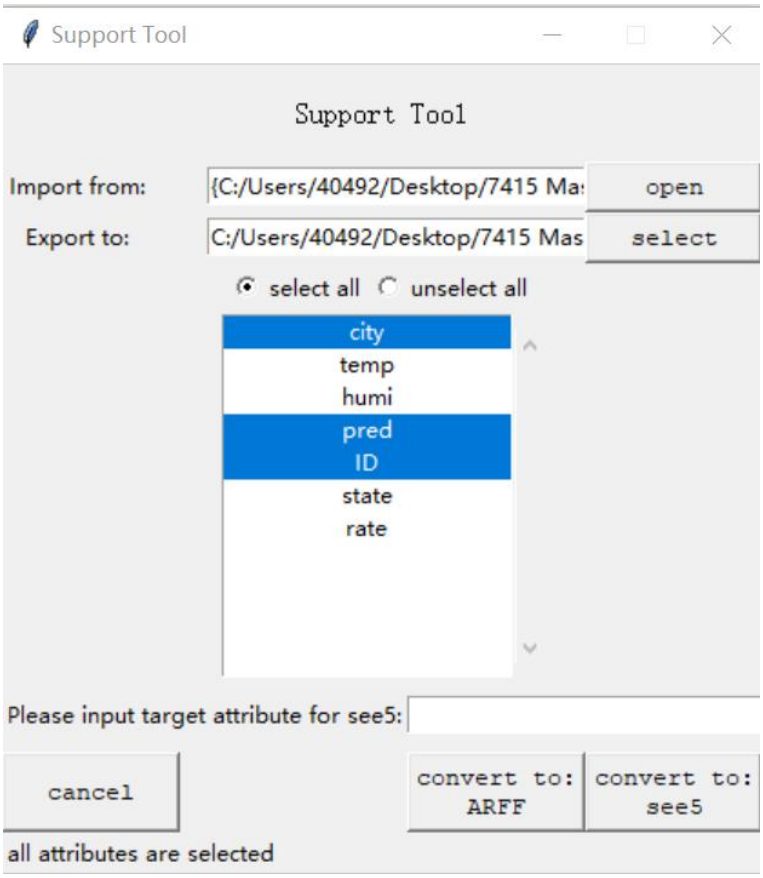
(2). Click “open” button, choose the csv file that you want to convert.



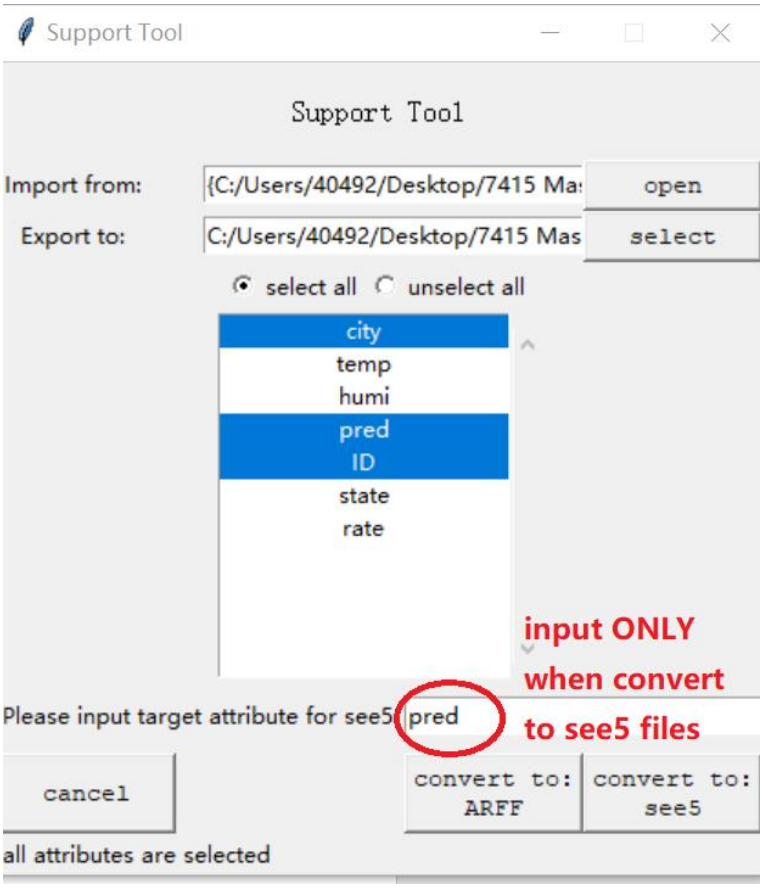
(3). Click “select” button, choose the directory where you want to save your new generated file in ARFF or see5 format.



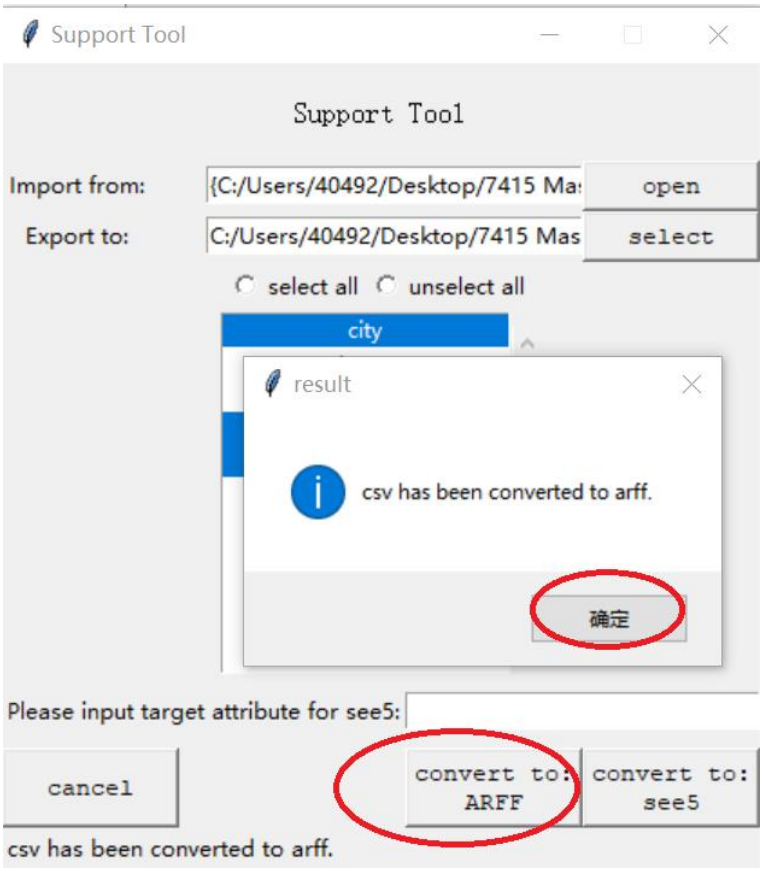
(4). Choose the attributes that you want to convert.



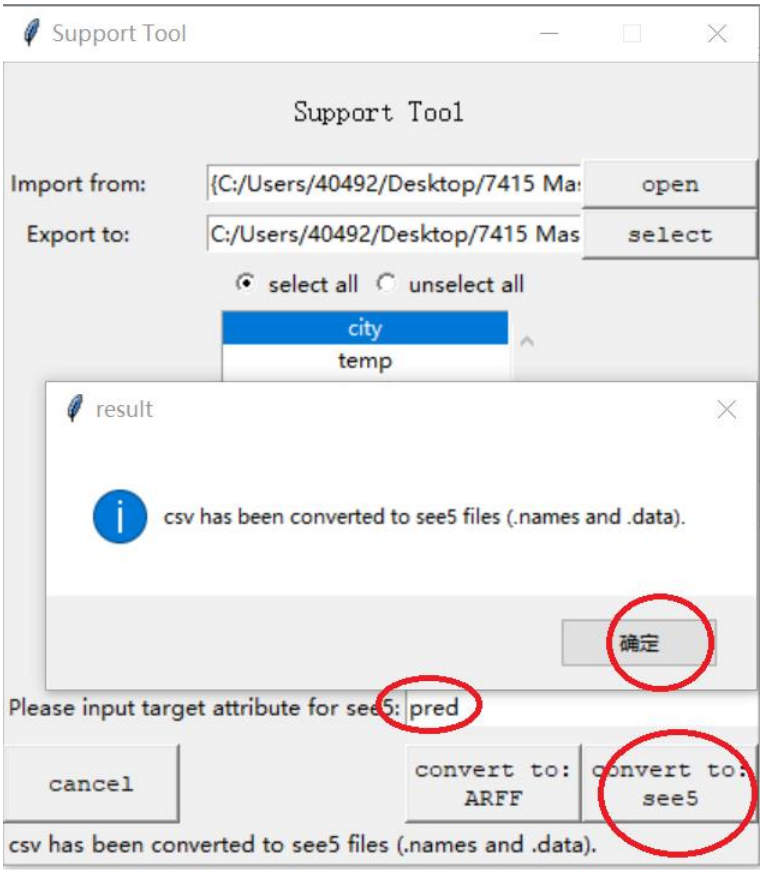
(5). Please input the target attribute in section (5) when you convert see5 files. Otherwise the target attribute will be null.



(6). Click “convert to ARFF” button or “convert to see5” button.

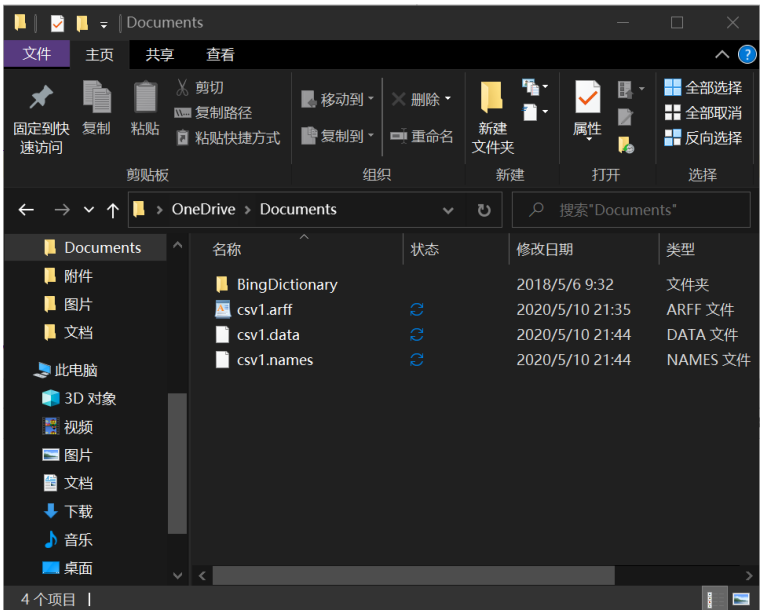


Or:

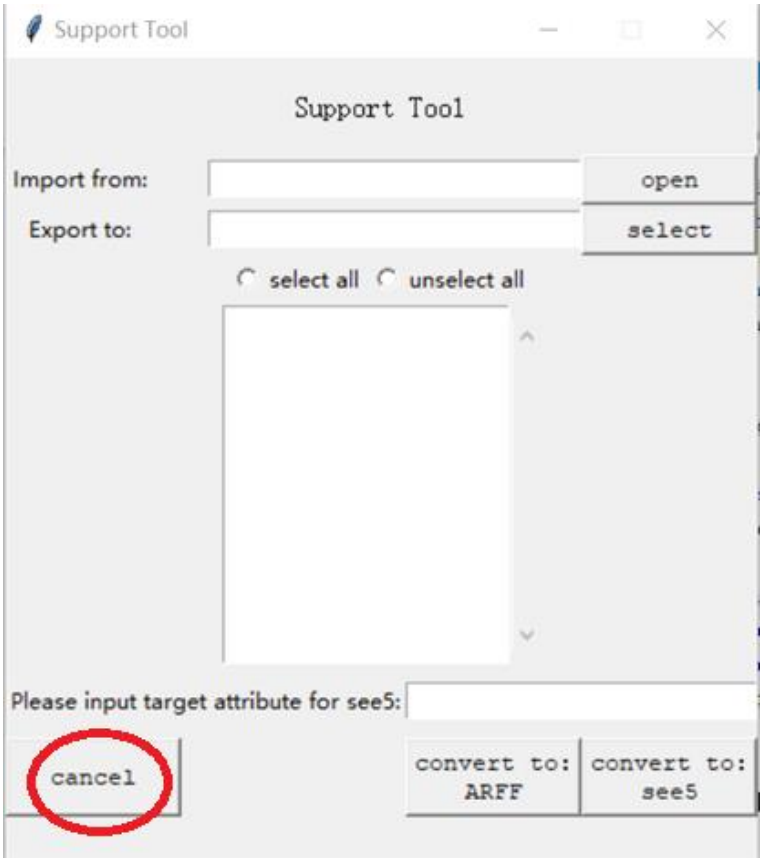


Notes: once the files are successfully converted, there is a result message pop out. And once you read and click yes, the window will be closed automatedly.

(7). And then, you will see the converted file(s) in the directory that you choose to save.

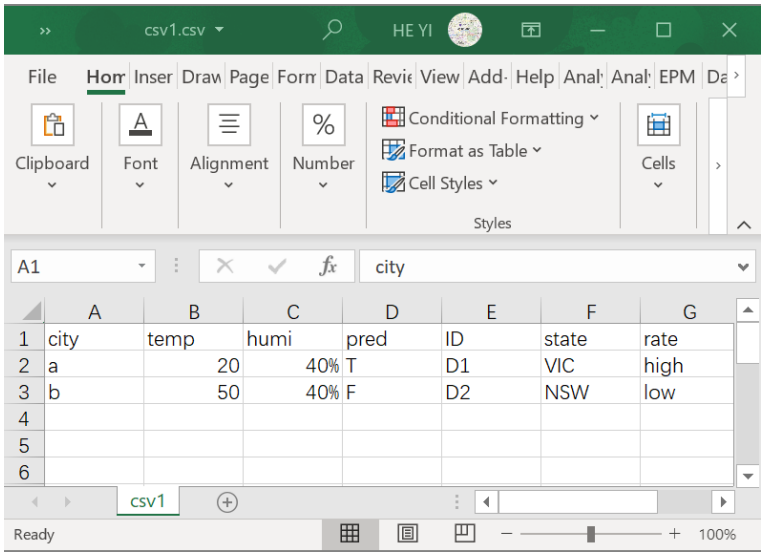


(8). Click “cancel” button to close the window and quit our converting product at anytime.



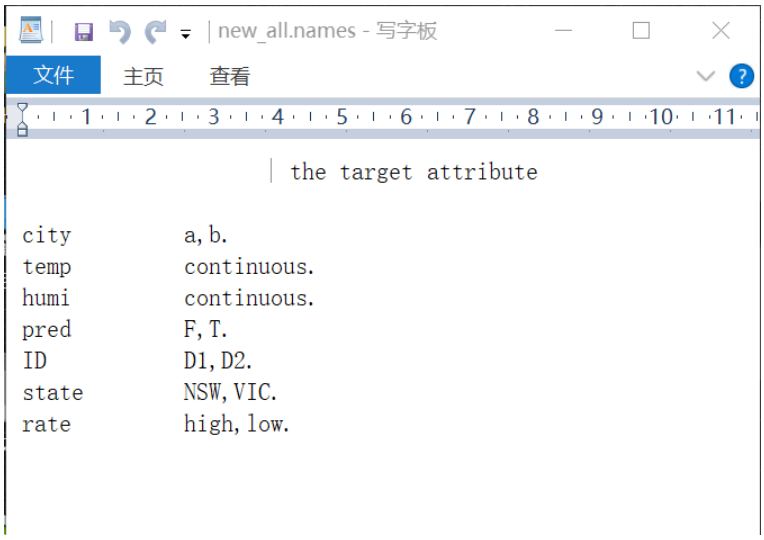
3. Sample of input and output screenshots.

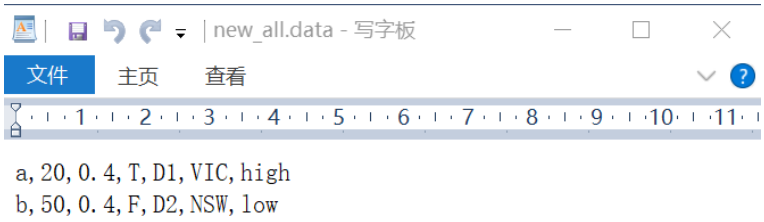
Input（csv1.csv）

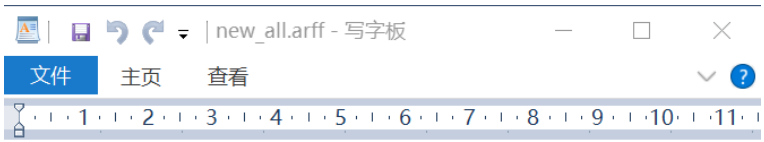


	A	B	C	D	E	F	G
1	city	temp	humi	pred	ID	state	rate
2	a	20	40%	T	D1	VIC	high
3	b	50	40%	F	D2	NSW	low
4							
5							
6							

Output1-for whole file (new_all.names, new_all.data and new_all.arff)





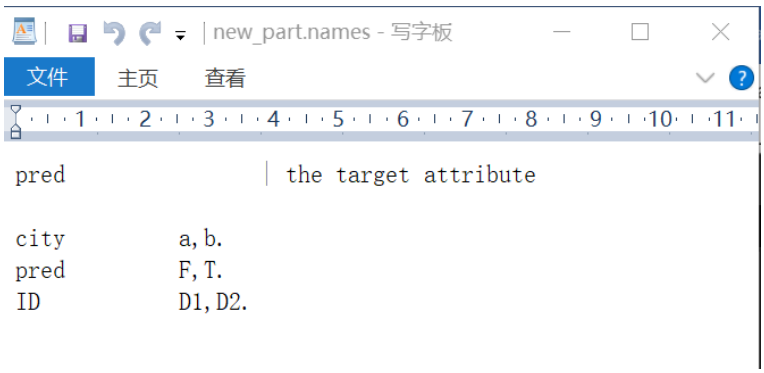



```
@relation 1591365386.1801586new

@attribute city {a,b,z,w}
@attribute temp numeric
@attribute humi numeric
@attribute pred {T,F}
@attribute ID {D1,D2}
@attribute state {VIC,NSW}
@attribute rate {high,low}
@attribute postcode numeric
@attribute int numeric

@data
a,20.0,0.4,T,D1,VIC,high,?,?
b,50.0,0.4,F,D2,NSW,low,?,?
z,20.2,0.4,T,D1,VIC,high,3350.0,30.0
w,50.0,0.4,F,D2,NSW,low,3000.0,30.0
```

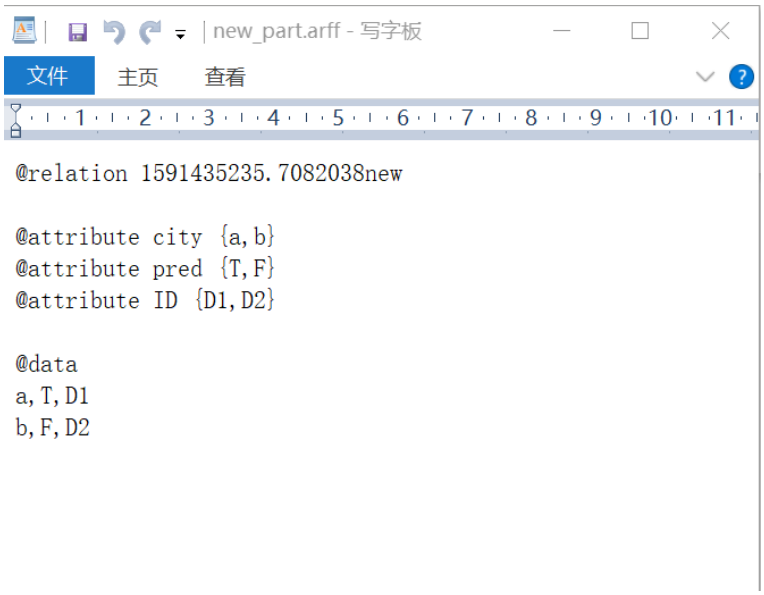
Output2-for partly converting (new_part.names, new_part.data and new_part.arff)




```
pred                | the target attribute

city          a,b.
pred          F,T.
ID            D1,D2.
```

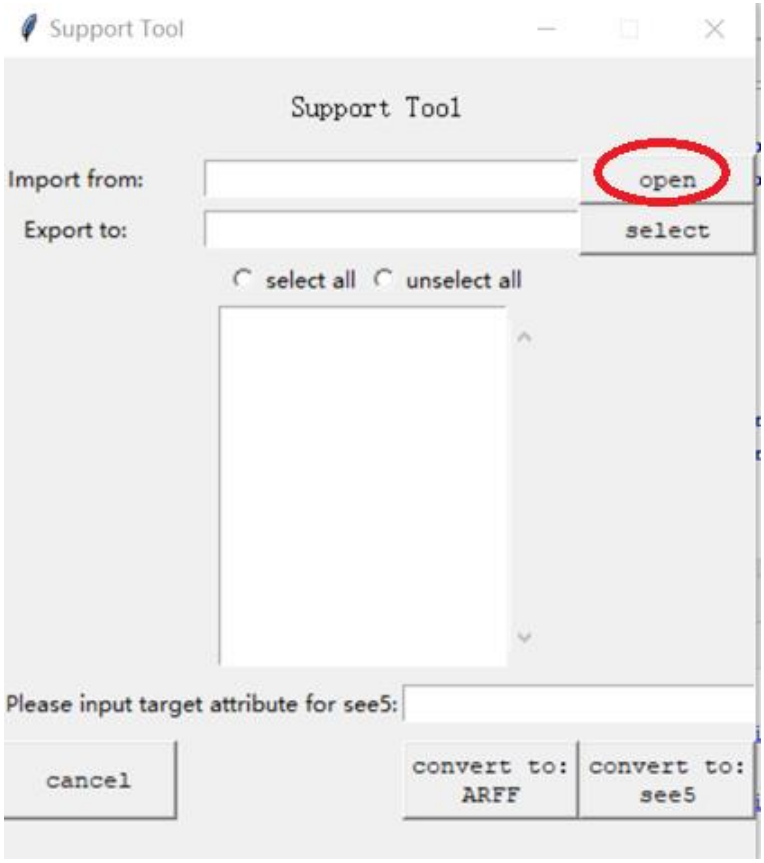

```
a,T,D1
b,F,D2
```

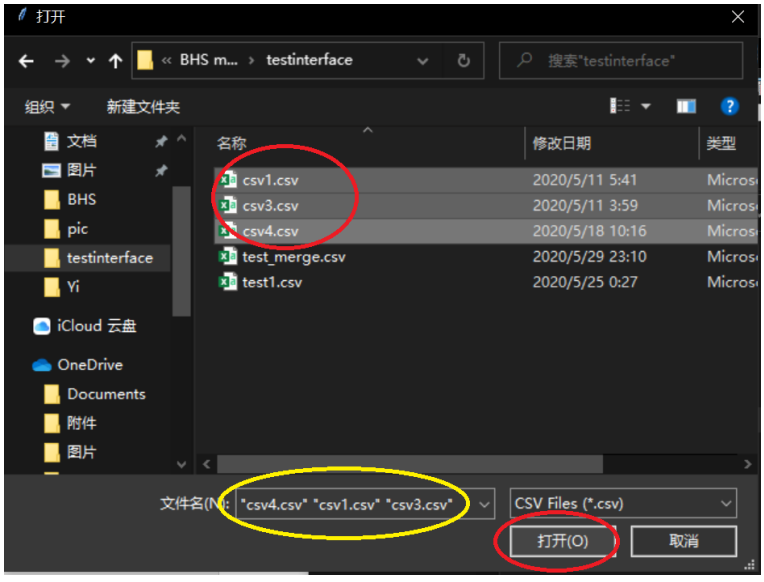

new_part.arff - 写字板
文件
主页
查看
@relation 1591435235.7082038new
@attribute city {a,b}
@attribute pred {T,F}
@attribute ID {D1,D2}
@data
a,T,D1
b,F,D2

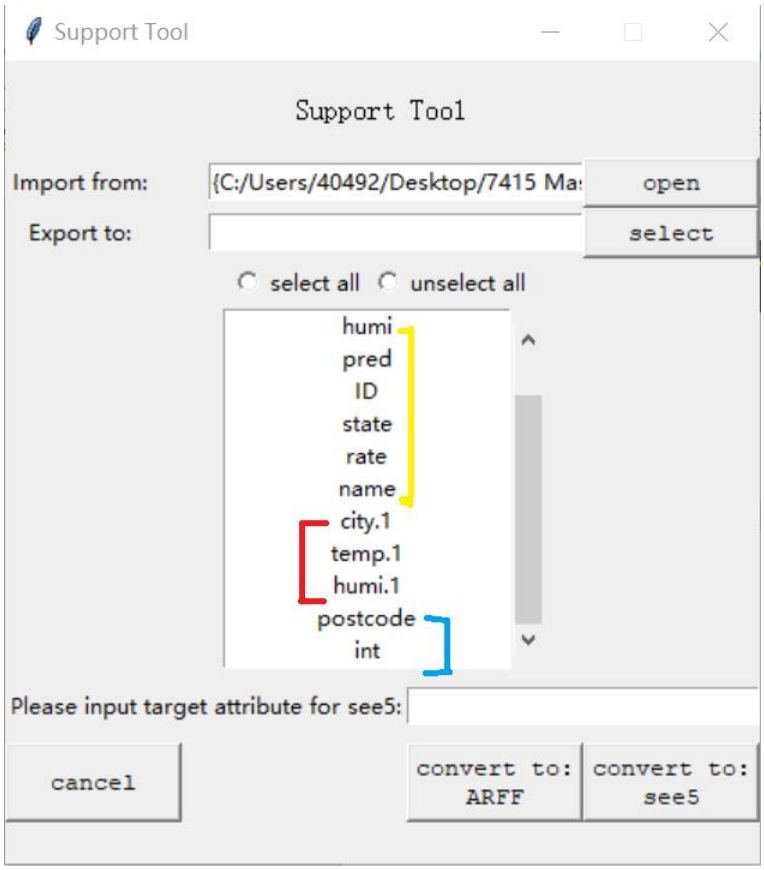
Multiple files conversion

Merging data

Our product can deal with multiple CSV files at the same time with data cleaning function. The only thing you need to do is select several CSV files while you click “open” button. And the merged data’s attributes names will be presented in the middle of the user interface (i.e. Listbox content will be updated according to your selection).





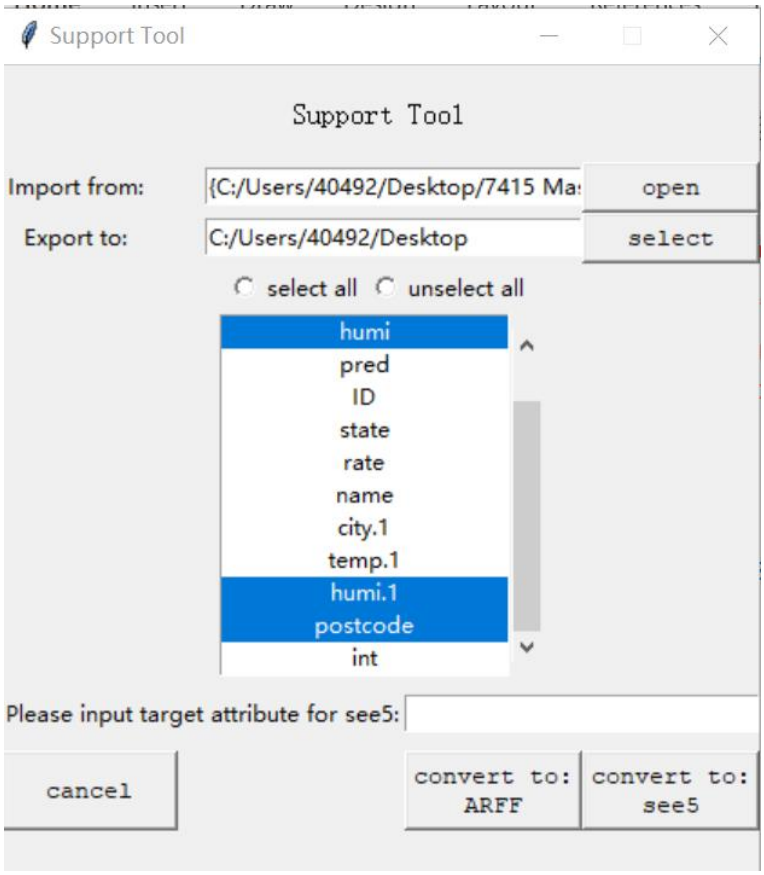


NOTE:

The yellow part of attributes are from the first file (csv1.csv), red from second (csv3.csv) and blue from third (csv4.csv) in this example. The same attributes will be merged and missing values will be filled with “?”. We cannot see the “?” in our interface because only attributes names are exhibited in the listbox.

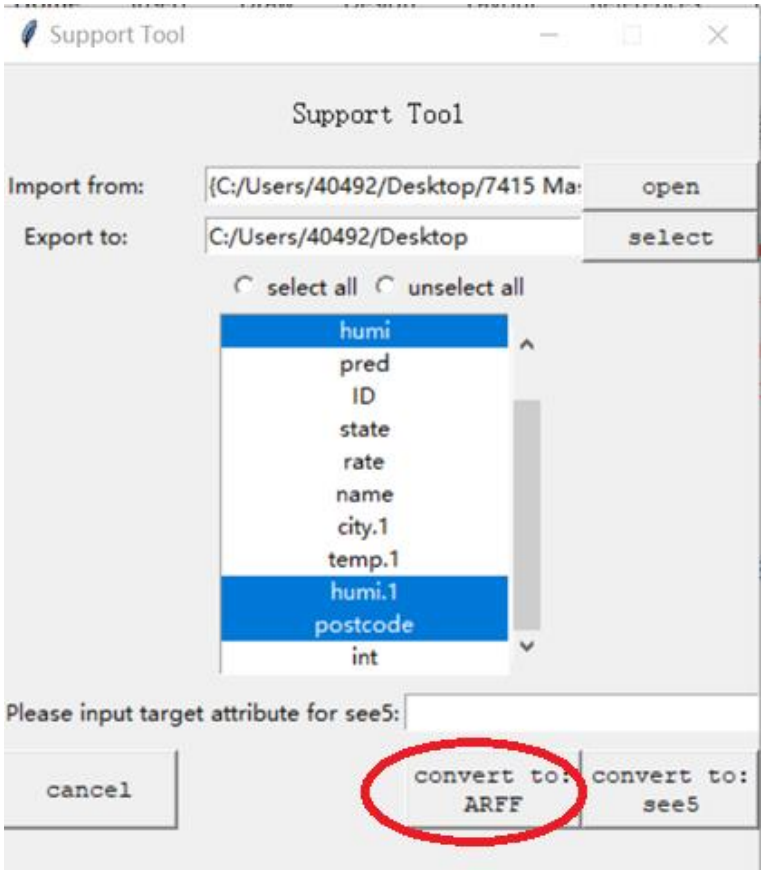
The “humi” and “humi.1” for instance in the above screenshot, are same attributes names after merge all the CSV files you select, but the values below the attributes are different. So we treat them as different attributes and rename the second “humi” as “humi.1”.

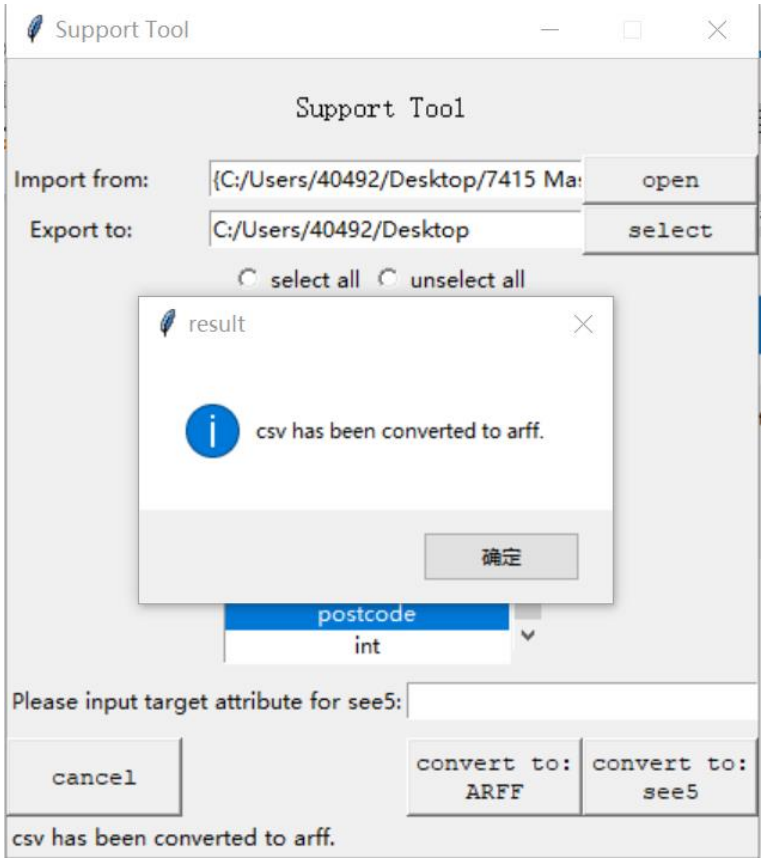
Partly conversion



Our product support partly conversion of the data. What the users need to do is just click the attributes you want to select via cursor and “Shift” keyboard.

And then click the format (ARFF or see5) you want to converted to. In this case, we use “Convert to: ARFF” as an example.

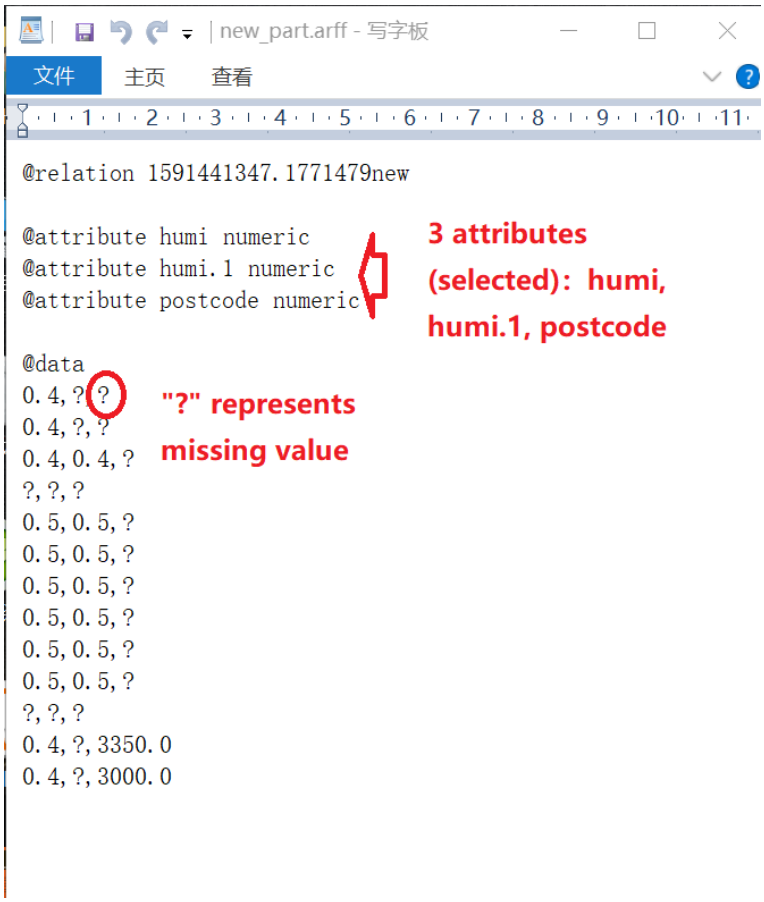

Support Tool
Support Tool
Import from:
{C:/Users/40492/Desktop/7415 Ma:
open
Export to:
C:/Users/40492/Desktop
select
select all
unselect all
humi
pred
ID
state
rate
name
city.1
temp.1
humi.1
postcode
int
Please input target attribute for see5:
cancel
convert to: ARFF
convert to: see5


Support Tool
Support Tool
Import from:
{C:/Users/40492/Desktop/7415 Ma:
open
Export to:
C:/Users/40492/Desktop
select
select all
unselect all
result
csv has been converted to arff.
确定
postcode
int
Please input target attribute for see5:
cancel
convert to: ARFF
convert to: see5
csv has been converted to arff.

Once the conversion is done, we will get a pop-up messagebox to show status. After you confirm the information (click “yes (确定)”), the .arff file will be generated and the user interface will be closed automatically.

Note: some of the button shows non-English letter due to the language of the users operating system. If you are using English version for your computer, you won’t see other languages.

In above instance, we select to save the file in desktop. So when we go back to the desktop, we find a ARFF file called “new_part.arff” on the desktop. That is the partly converted ARFF file. Open it and we can see the content as below.



Frequently Asked Questions

1. Do I need coding skills to use this product?

No. You don't need equipped with any coding skills for the use of our product. What you need to do is just follow the instructions and click on the user interface and then you can get what you want, such as ARFF file for Weka or NAME and DATA files for see5.

2. Do I need to install any packages to make the tool work?

We provide users with an easy access to our product that if they can run the execution file (i.e. "supporttool.exe") directly on the Microsoft windows operating system without installing anything in PCs. However, the traditional approach need to installation before running the script.

3. How the data is cleaned compared to the original CSV files?

Once you select multiple CSV files via the "open" button, our product will check the attributes names (column names – the first rows of the CSV files) and compare them and the data in the same column. If the attribute name and the data are same, we recognize it as a coincide attribute, so we will just keep one unique column in the output. Meanwhile, if there are duplicated rows in the merged data, we will only leave one unique row. Missing values will be filled as a "?" instead of a blank cell in the final output files.

4. Is the output readable by Weka and see5?

Yes. All the outputs have past all the testing. Although in some parts it's not exactly same as the analytics software expected way to deal with special marks, such as percentage, but our product provides fabulous functions which enable the output files works better in the future use by Weka and see5.

For instance, we convert percentage values into float data, so that all the percentage data won't be recognized as a label but a number. So it's more practical when doing analyses in Weka and see5.

Another example is we don't change the order of the attributes, so users could easily find the attributes according to the files they open.

The final tiny difference for data cleaning when opening multiple CSV files, integer data will be transformed into float data with 0 after decimal point but it won't influence the precision of the values. However, one file conversion isn't been influenced. That is, the output of one file conversion will keep integer as integer data after conversion.

5. Is the product can only convert for once?

Yes. Once the conversion is done, the user interface will be closed automatically. If you want to continue converting other files separately, you need to reopen our product.

Meanwhile, you need to rename the generated files' name before continuing for this product will override the file if there's an existing file which has the same name. So if you want to keep convert different files separately you need to rename or restore your file manually.

We can also provide tool which can keep the same name as the CSV files when you only convert just one file, that is, without data cleaning and merging function. For more detail about our products, please send email to: yihe@students.federation.edu.au.

Copyright